Managing complex engineering projects: What can we learn from the evolving digital footprint?

Abstract

The challenges of managing large complex engineering projects, such as those involving the design of infrastructure, aerospace and industrial systems; are widely acknowledged. While there exists a relatively mature set of project management tools and methods many of today's projects overrun in terms of both time and cost. Existing literature attributes these overruns to factors such as unforeseen dependencies, a lack of understanding, late changes, poor communication, limited resource availability (inc. personnel), incomplete data and aspects of culture and planning. Fundamental to overcoming these factors belies the challenge of how such potentially critical management information can be provided, and done so in a cost effective manner. Motivated by this challenge, recent research work has demonstrated how such management information can be automatically generated from the evolving digital footprint of an engineering project covering a broad area of methods and data sources. In contrast to existing work that has reported the generation, verification and application of methods for generating management information from different data sources, this paper reviews all the reported methods to appraise the scope of management information that can be automatically generated from the digital footprint. In so doing the paper presents a reference model for generation of managerial information from the digital footprint, an appraisal of 27 methods, and critical reflection of the scope and generalisability of data-driven project management methods. Key findings from the characterisation include the role of email in providing insights into potential issues, the role of computer models in automatically eliciting process and product dependencies, and the role of project documentation in assessing project norms. The critical reflection, raises issues such as privacy, highlights enabling technologies and the opportunity for new business intelligence services based on real-time analysis of digital footprints.

Keywords: Big Data, Project Management, Business Intelligence, Knowledge Workers

1. Introduction

Advances in science and technology, societal need (e.g. population and quality of life), and inter-connectedness of socio-technical systems combined with mankind's aspirations to explore and explain the world round us, demand that today's engineering projects, no matter whether hardware or software, deal directly or indirectly with a combination of scale, complexity and inter-dependency. Correspondingly, engineering projects are themselves large (inc. value, resources, and people) and complex (interdisciplinary, distributed, and long-life) making effective management to time, to cost, and to quality highly challenging even for the most experienced organisations and project teams. High profile examples of overruns in terms of time and cost include: the Airbus A380 - 2 year delay and £1.9bn overrun cost [26]; the Boeing Dreamliner - 3 years late and an estimated cost increase from $6bn to $32bn [45]; Sydney Opera House - 1,357 percent overspent and 10 years late [50]; and the Olkiluoto Nuclear Power Plant in Finland which remains incomplete and almost 300 percent over budget [51]. Further, there are a number of high profile projects that have been withdrawn or closed down early such as the UK's NHS Patient Record System project which was cancelled accruing a spend of £10bn [43]. Such cases have been examined in detail by both industry and academia with failures attributed to exceptions, scope creep, communications, resource availability and aspects of culture and planning [32, 33, 49, 28, 34, 4, 6]. While in some cases a single trigger event can lead to failure, more often than not failure is due to the culmination and combination of a complex intertwined set of issues. Correspondingly, the monitoring of simple project measures alone (e.g. task completion and resource consumption) are insufficient to unravel and mitigate the many compounding factors. Richer management information is required, supplementing outcome or progress data with detailed information on the state, issues, and outcomes of activities that previously or currently being undertaken. In the context of engineering projects, these activities include but are not limited to: design work [3]; CAD tasks [5]; email activity [48]; information seeking and sharing [35]; and, collaborative work [27]. The level of detail of information necessary to fulfil this requirement and the associated resources and effort necessary to generate the information pose an insurmountable barrier, not to mention the fact that if manually generated the management information would likely lag by a time period that would render it out-of-date.

In today's digitally enabled workplaces many sub-activities and tasks are undertaken via a digital tool/tool chain, e.g. email communications, CAD design, simulation, and reporting via technical documents or presentations. Correspondingly, the major part of project activities will be undertaken using and/or reported via a digital tool that in turn generates a digital file. These digital files are created and evolved in almost real-time as work is performed, and can thus be considered to provide a snapshot and history of activity i.e. that which has, and is, being undertaken. Where an engineering project is considered, these digital files collectively form a digital footprint of the project that evolves and mirrors, with minimal lag, the project activity. Following approaches of data-mining, it follows that analysis of this near-real-time gives potential to automatically provide management information. Since 2013, a large number of methods (20+) have been demonstrated to automatically generate management information from the digital footprint of engineering projects. These previously reported methods form the major of the data used in this paper. The individual methods are detailed in Section 3 and Tables 4 and 5.

In contrast to existing work that reports the generation, verification and application of individual methods [47, 21, 36], this paper reviews the various methods collectively with the aim of appraising the scope of management information that can be automatically generated from the digital footprint, thereby addressing the question of *'What* *can we learn from the evolving digital footprint?'*. The paper begins with a précis of current project management tool-sets, the nature of the management information they provide, and the difficulties in provision of richer management information (particularly in regard to resources, effort and timeliness). The paper then presents a reference model that describes the various methods that have been proposed for generating types of management information from the digital footprint. The model provides the framing for the cataloguing and subsequent appraisal of the scope of management information that can be generated from twenty-seven catalogued methods (See Tables 4 and 5). The paper concludes with a reflection on scope and generalisability, the implications for data-driven project management, and the potential barriers to industrial uptake.

2. Existing project management toolsets and management information provision

Current support for the management of engineering projects can be considered to span four areas: project management methodology; project management software; project management tools; and, project performance measures (e.g. Gross Profit, ROI, Earned Value, Cost, Variance, Customer Satisfaction [8]). In most cases a single project methodology and single software suite will be adopted with a range of project management tools and performance measures employed. Examples are given in Table 1. In the majority of cases, the selection of these elements will be based on previous projects, recommended practices or guidelines and/or company procedures. Further, certain measures may not be compatible with certain methodologies. For example, Earned Value Management (EVM) requires quantification of a project plan, hence it is generally considered to be unsuitable for discovery-driven or Agile projects where it may not be possible to fully plan projects sufficiently far in advance. In addition to the methodologies, software, tools and measures, there exist a number of project management standards. These include ISO10006 / ISO 21500 [20], PMBoK [2] and APMBoK [1] which were developed to explain and provide guidance on the core principles and good practice(s) in project management.

This paper aims not to critique each methodology, tool, or measure but to consider collectively what they represent in terms of management information provision. While the benefits of such methodologies, tools, methods and standards are widely acknowledged, their primary application lies in planning and cost estimation (upfront) and measurement of cost and progress during a project. Correspondingly, and given also their reliance on analogues derived from idealised/generalised planning and historical data, e.g. for resource consumption or schedule, few tools and methods provide detailed information on the state, issues, and outcomes of activities that have or are currently being undertaken. As stated in the previous section and in the context of engineering projects, these activities include for example: design work [3]; CAD tasks [5]; Email activity [48]; information seeking and sharing [35]; and, collaborative work [27]. Where methodologies do provide some understanding this is either at an abstracted or aggregate level, such as 50% of design review complete, or via supplementary organised activities such as scrums or huddles that underpin agile methods. While beneficial these manual methods can become unwieldy for large teams, impractical for distributed teams, and are limited in scope due to their time and resource commitment. Further, the workload associated with generation of information through such means will likely create a time-lag, with consequent impact on ability for managers intervene effectively.

In the case of routine or well-understood engineering projects the relationships between activities and the state, issues, and outcomes of activities may be easily observable and/or well understood (repeatable). For example, the process of designing a structural member for a high performance car might typically follow an accepted and logical process. This process would involve: continuous communication with another subsystem team; initial CAD

Project Management support	Examples and definitions
Methodologies	Waterfall, Critical Path Method (CPM), Critical Chain Project Management (CCPM) Agile, Scrum, Kanban, Extreme Programming (XP), Adaptive Project Framework (APF)
Software	Examples include Mavenlink [19] and LiquidPlanner [18]. Typical provide support for: collaboration, idea management, portfolio management, requirements management, resource management, task management, testing and QA management.
Tools	Gantt chart, Logic Network, PERT chart, risk registers, Product Breakdown Structure (PDS) and Work Breakdown Structure (WBS), and project networks.
Measures	Measures generally relate to aspects of project accounting and cost management/monitoring, schedule and productivity as well as supplier performance [8]. More advanced techniques such as Earned Value Management (EVM) and Earned Schedule (ES) can be employed to monitor variances with respect to schedule (plan).

Table 1: Common methodologies, software tools and measures for engineering project management

work; review of interfaces; detailed CAD work; the consideration of existing or standard parts; virtual testing; refinement; final review; reporting; approval; generation of tool paths; manufacture; inspection and sign-off. For the purpose of project management this might be incorporated into an overall sub-system design activity (e.g. chassis) or a separate activity that is broken down into design, development and production activities. Consequently, experienced project managers can bring to bear their own knowledge of this relation while managing the project. In contrast, for complex and/or one-off engineering projects the initial programme definition is not known *a priori*. As such, it is not possible to form a complete definition of activities and their interrelationships, let alone comprehending the intricate dependencies between progress, issues, and outcomes of lower-level tasks. Consequently, a learning approach and mind set are required [7] – a requirement that is arguably a prerequisite for the adoption of agile approaches. Fundamental for this learning is the ability to access and interrogate information about the state of a project in such a manner as to provide evidence for interventions, and detailed (richer) management information and to do so in real-time or as close to real-time as possible, and with minimal manual input. It is contended in the aforementioned précis that current tool sets do not presently provide this capability and that new supplementary tools/methods are required that can provide such management information in a more automated manner. Further, where complex engineering projects are considered such rich and automatically generated information is imperative for delivering t time, cost and quality. In the previous section the opportunity to mine the evolving digital footprint of an engineering project is proposed and developed further in the following section.

2.1. The digital footprint of engineering projects

As previously stated the digital footprint of an engineering project is an evolving record of engineering work that embodies what has been done and is currently being done. The digital footprint comprises all the digital data that is generated by members of the project, towards the intended project outcomes. Due to continued reliance on digital tools, the digital footprint increases in size as the project progresses. It comprises of a variety of file types, such as the sixteen classes set out by [15] which includes the common digital data shown in Figure 1. For large engineering projects that involve many organisations and span many years this digital footprint can comprise many

100,000s of files and communications that, with the exception of communications, will each have been modified many 10s or 100s of times. Further, and almost without exception, all forms of engineering activity generate, in almost real-time, a digital shadow, even today's engineering logbooks [29]. The comprehensive and almost real-time nature of the evolution of the digital footprint presents an opportunity to computationally mine the digital footprint in order to continuously monitor and appraise the project. Such an approach is analogous to data-driven methods of condition monitoring that are maturing within automotive and aerospace engineering. A commonly cited example is that of Integrated Vehicle Health Monitoring (IVHM) [22] that has been applied to vehicles and high value assets such as wind turbines and gas turbines [44]. The principles of IVHM and how they can be applied to the digital footprint of an engineering projects is discussed in [41] and is not thus not developed in detail in this paper. Rather the focus is on the various data-driven methods themselves. In the context of IVHM and digital footprints these methods take the raw data (sensor data or changes to digital files respectively) and process the data (manipulate or analyse) in order to provide information about the state of the asset or project respectively.

Based on these principles and the need stated earlier in this section, the Authors have undertaken a large number of studies into the understanding and insights that can be generated from analysis of types of digital asset (file) including CAD models [14], technical reports [37], email [47], presentations and social media [12]. These studies and their corresponding publications are set out in Tables 4 and 5 and Figure 5. For the purpose of this paper and, in particular, to catalogue and characterise the various methods, file types are dealt with under five groupings: email, technical reports, computer models, project documentation, and all (the entire collection of files).

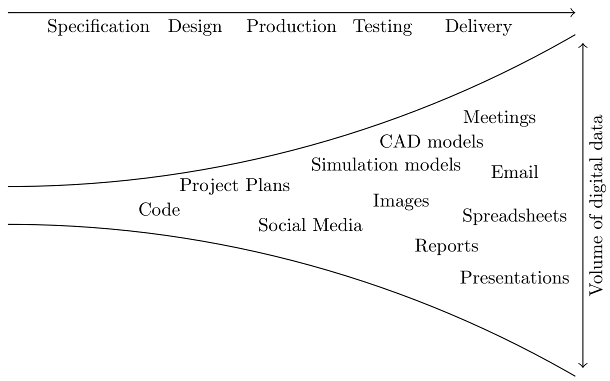


Figure 1: Growth in the volume of the digital footprint during the typical stages of an engineering project

3. Research approach, reference model and cataloguing

As previously discussed, the aim of this paper is to review, catalogue, and appraise the range of reported methods, towards the ultimate aim of addressing the question of *'What can we learn from the evolving digital footprint?'*.

In order to achieve this, previously reported methods - as defined in Section 2.1 - are reviewed with the objective of eliciting a reference model that can be applied to catalogue and characterise each method. Once established the reference model is used to catalogue existing and recently developed methods, following which the methods are appraised by virtue of two complementary perspectives: i) classes of management information against types of digital asset; and ii) the interpretive power of the management information with respect to engineering projects. The overall research approach is illustrated in Figure 2.

3.1. Reference model

Based on a review and characterisation of the embodied processes within the reported methods, c.f. Tables 4 and 5, a generalised form or reference model can be developed. The reference model is depicted in Figure 3 and provides the basis for cataloguing the various methods. The reference model has been developed through consideration of the Open System Architecture for Condition-Based Maintenance [30] and decomposition of the reported methods

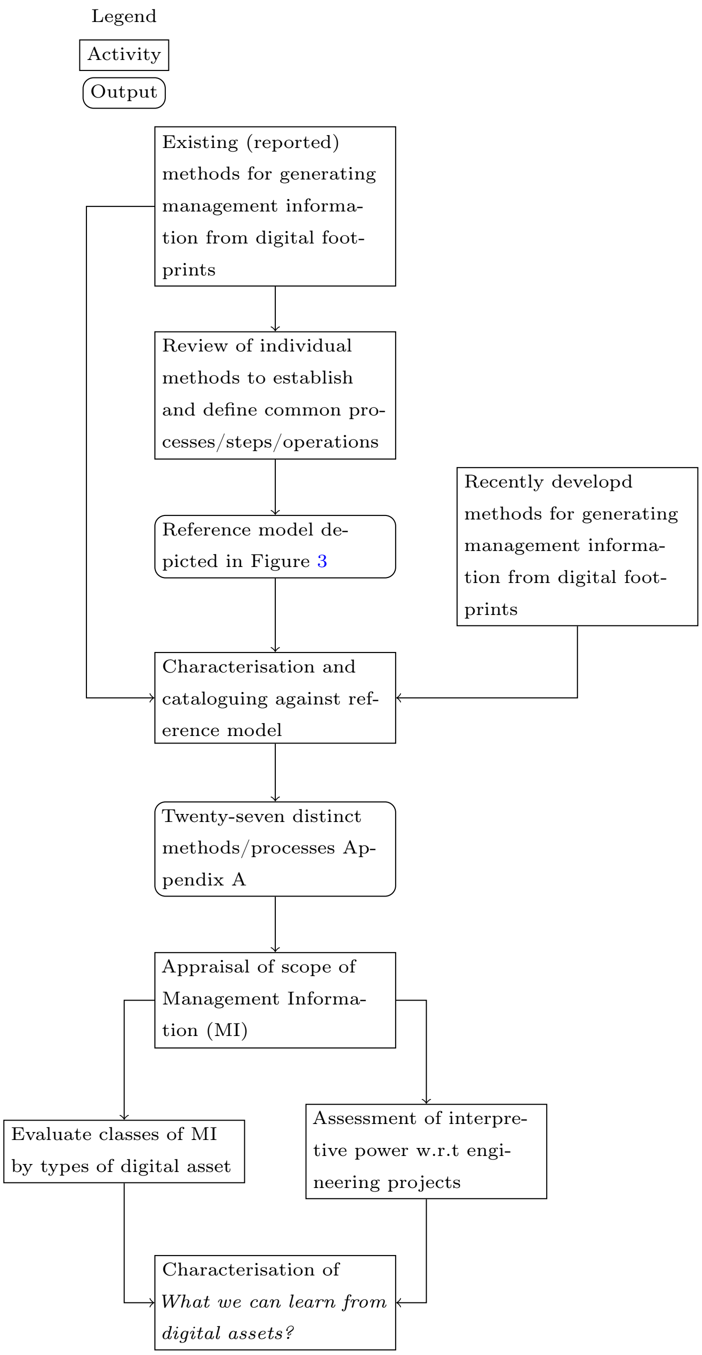


Figure 2: Cataloguing, characterising and appraisal of methods

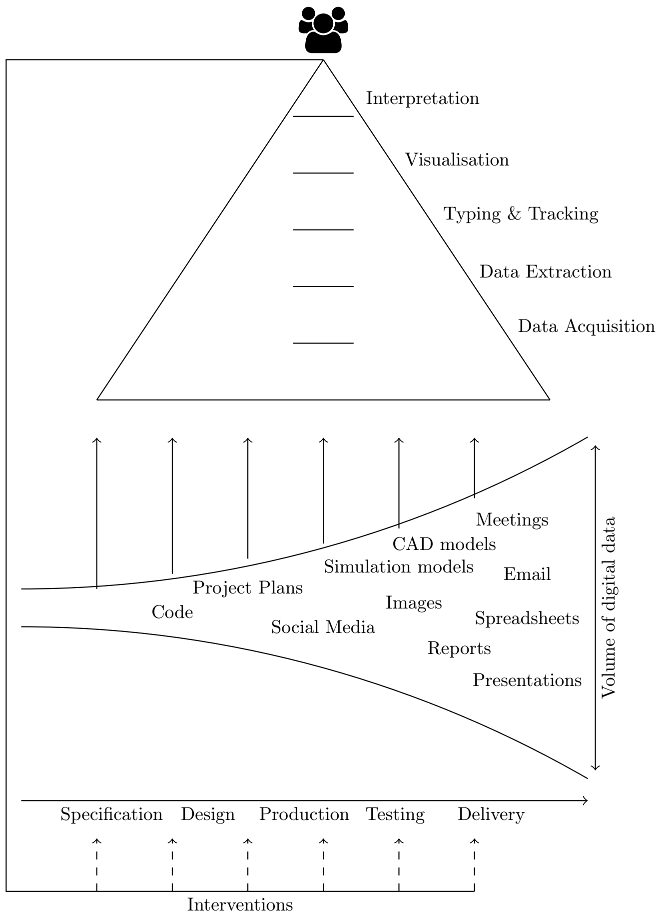


Figure 3: Reference model for methods that generate management information from digital assets

(Tables 4 and 5) and comprises five-stages; four stages that involve data-information processing: data acquisition, data extraction, typing and tracking, and visualisation; and a fifth user-processing stage entailing interpretation by a project manager and/or stakeholder. This latter stage reflects that such data-driven models are intended to supplement existing methodologies and tools (c.f. Section 2), providing detailed information in support of managerial decision-making and intervention. The four data-information processing stages are now discussed in detail in order to provide the structure and framing necessary for repeatable curation/cataloguing of methods.

- Data acquisition is the process of capturing changes in the digital footprint. This includes the current state and instances of past states, where a state includes new assets (files) and the changes to existing assets. For the purpose of data acquisition, some components of the digital footprint evolve sequentially and cumulatively, such as email, while others evolve continuously as they are modified, such as a technical report or CAD model. Correspondingly, instances of the digital footprint must be captured at appropriate time intervals e.g. hours or parts thereof.

- Data extraction is the process of opening, identifying and copying specific data from within an asset(s) and then preparation of the data for further processing.

- Typing and tracking are the two fundamental operations that are applied to the data in order to provide the basis for the generation of management information. Typing is the process of adding contextual meaning to the data and involves classifying individual data elements, occurrences of data types, events in the life of the data, clusters of data, and patterns of, and within the data with respect to a particular perspective. Examples include typing language by sentiment, by purpose, by function in the workflow, or by role in the project. The Authors note that the spectrum of 'types' is potentially vast. Tracking is the process of measuring the changes to data, the occurrences within data, the relations between data and/or typed data.

- Visualisation is the process of constructing a representation of typed and tracked data over time, for the purpose of interpretation. Dependent on nature of data, this may be presented with respect to the absolute and relative changes over time / time period, and framed with respect to either the process (overall project plan), people (team) or the product (system).

3.2. Cataloguing new and existing methods

As previously stated in Section 3, the range of possible methods is potentially vast - dependent on the combination of available data and spectrum of 'types' (c.f. typing and tracking in Section 3.1). Consequently, it is neither feasible nor the aim of this paper to establish the set of all possible methods. Rather, we apply the reference model in order to characterise and catalogue the set of methods that have been derived practically by the Authors from a variety of real engineering projects and the needs of real engineering teams. These reported methods and recently developed methods have been developed and verified across a large number of complex engineering projects over a four year period. Complex projects are considered to be those of scale (value, size and number of project members), and involving interdisciplinarity, novelty and criticality (i.e high value/safety critical systems such as aerospace). The engineering projects studied are summarised in Table 2 and took place between September 2013 and July 2017. Projects ranged in duration from weeks to years, in personnel from 10 to 500+, and covered the domains of aerospace, automotive, software and industrial systems. Thereby providing a x-domain and representative cross-section of engineering projects.

For the purpose of developing the various methods, the studies were undertaken on both live and completed engineering projects depending on the company, availability of data and access to personnel. In both cases (live and completed) project members and/or stakeholders were consulted in the development of the methods, the management information generated, and their interpretation within the context of the engineering projects under study. For the purpose of developing and verifying the methods, a series of user studies and workshops were also undertaken and are reported in [39, 31]. In combination with this, an extensive literature review of features of interest of engineering projects was undertaken and can be accessed in [38]. Through these perspectives (user studies, workshops, and proposed methods) combined with inspection of the available data (c.f. Table 2) and feedback from project stakeholders, a set of twenty-seven methods for the generation of management information has been established and is given in Tables 4 and 5. These represent a comprehensive (yet knowingly non-exhaustive) set of distinct, practical, and observable methods given the composition of engineering projects studied, the data sets provided, and the known factors that impact on project success. For the purpose of

Study	Sector/Type	No. of Personnel	Duration	Digital Footprint	No. of Project	Live or Completed
1.	Aerospace engineering	10+	1 week to 3 months	Workflow including all reports and formal communications	100+	Completed and live
2.	Systems engineering	500+	3 years	E-mail	1	Completed
3.	CAD/CAM software	15-40	6 month cycles	Workflow, version control and documentation	3 cycles	Live
4.	E-commerce platform and development	6-15	6-12 months	Documents and e-mail	5	Completed
5.	Formula Student	30-40	2 years	Reports, social media, CAD models, code and simulation	3	Completed and live
6.	Manufacturing process/systems	15-30	3-12 months	Reports and presentations	6	Completed and live

Table 2: Characteristics of the engineering projects studied

the studies undertaken (Table 2), the management information generated from the methods were implemented using a variety of common information visualisation techniques including time-series, network graphs, strategic diagrams and tree maps. In order to illustrate the methods and management information, sanitised visualisations for a range of the management information (one for each class of management information - c.f. section Section 4.1) are given in Figure 4 and summarised. Figure 4 also includes references to further information and, in the majority of cases, the underlying research that informed, verified, and validated the specific method(s).

4. Appraisal and characterisation of the scope of management information

For the purpose of characterisation and appraisal of the scope of management information, two complementary perspectives are considered. The first is an assessment of different classes of management information against types of digital asset. The second is an appraisal of the interpretive power of the resultant management information with respect to engineering projects.

4.1. Classes of management information

Table 3 presents the twenty-seven methods by first grouping them with respect to the management information they provide and then associating them against the type(s) of digital asset from which they are derived. For the purpose of the analysis a check mark is used to denote that one or more of the methods within the class can be derived from the type of digital asset. For example, in the case of *topic* all five methods are derived from e-mail which the two *product* methods may be derived from a number of types of digital asset. Table 3 also details the total number of reported methods in each class. For the purpose of this paper, the methods have been classified into nine groups based on the management information that they generate.

1. Communications - management information that relates to or represents features of the communication within a project team, including content and transmission.
2. Conformance - management information that relates to or represents conformance or compliance of project work with respect to a predetermined or formalised set or procedures, targets, or structure.
3. Dependencies - management information that relates to or represents relationships between aspects of the project, people, processes, or product.
4. Engineering effort - management information that relates to or represents the level of focus or effort given to engineering activities, tools, methods, or principles.
5. Product - management information that relates to or represents functional, behavioural, or structural

aspects of the product being designed.

6. Skills and competencies - management information that relates to the overall capacity or capability of the project team.
7. Time - management information that relates to the relative completion and / or likely completion date of project tasks and/or engineering activities.
8. Topic - management information that relates to, or represents the key concepts, themes and tasks that are receiving or require attention.
9. Workflow - management information that relates to or represents the relative order/sequence of activities or tasks.

The category and contents of these nine classes were developed through consultation within the research team and industrial collaborators. The nine classes were considered to be distinct based on the management information they provide and their interpretation but are not mutually exclusive. That is, they could not be easily combined into a single class but, for example, methods and management information in the 'time' class could also be classed as 'product'.

From inspection of the classes against the five types of digital asset (email, reports, models, project documentation, and all files) the following observations can be made:

1. Topic-based management information is derived almost exclusively from email. While it is possible to develop these from reports the relative lag in derivation of the information means that potential utility is compromised. That is, a report is generally post-fact and hence is more likely to incorporate past topics and consideration thereof, rather than current or emerging topics.
2. Management information relating to dependencies, product, and engineering effort are primarily derived from the engineering specific tools such as CAD (Computer Aided Design). This is because such tools embody the technical definition of the product or system being engineered, and correspondingly either implicitly or explicitly capture the internal relations within the product or system and reflect the stage of development (maturity). It is noted by the Authors, that email could also be used to develop management information relevant to this category.
3. Management information relating to effort are exclusively derived from those types of digital asset which, in the context of engineering projects, directly represent or explicitly involve technical engineering work (effort). These include technical reports and models. While assets such as email may discuss engineering effort they are not a direct measure of effort itself.
4. In contrast to effort-based management information, and on initial inspection perhaps a little surprising, is the fact that skills and competency related management information can be derived more fully from email than models. This is because sustained contribution to a particular technical discussion can be considered to be a clear indicator of ability to contribute and thus knowledge, skill, or competency. In contrast, while using a modelling tool might indicate proficiency with the tool, it does not afford a high fidelity indicator of level of competence or knowledge. Related to the rationale for using email to inform skills and competency information, technical reports are generally reviewed and approved, indicating a high level of competence of the approver in addition to author (often a senior member of staff e.g. Chief Engineer) and implying their validity as a data source.
5. Time-based management information is largely derived from the digital assets that form or contribute directly to the project deliverables. In the case of engineering projects, these are typically computer models and technical reports which reflect directly the technical definition of the product or system and

fulfil project deliverables respectively.

6. Workflow-based management information are not unsurprisingly derived from reports and stage-gate related activities. The former may be deliverables and denote the completion of activities or phases of the project e.g. review and approval, while the latter are explicit tasks associated with completion of a project phase or key activity.

From this categorisation and appraisal it is possible to gain some insights into the nature (class) of management information that can be derived from each type of digital asset given the content of the source and its application/role in the context of engineering projects.

4.2. Interpretive power of the management information

A key tenet in the derivation of the classes of methods and management information is their distinct interpretations in the context of engineering projects. For the studies reviewed in this paper, the assessment of interpretation of the management information was elicited through workshops, discussion and user studies. The interpretations for each class are summarised in the right hand column of Table 3. In contrast to Section 4.1 which considers the classes of management information that can be derived from each type of digital asset, this section appraises and attempts to characterise the interpretive power of the management information that can be derived from each type of digital asset.

Based on the classification and summaries of interpretive power the following observations can be made:

- Email is the primary asset necessary for generating management information concerning potential issues and holistic understanding of the team (e.g. level of collaboration and sentiment). The reason for this is that electronic communications are the most likely class of digital asset to reflect emerging issues and real-time triage of issues. Further, due to the networked nature of discussions and the conversational style adopted by many users of email, this class of asset most strongly embodies strength of feeling, opinions and emotion(s) between groups. Email therefore offers the greatest opportunity for the generation of understanding and insights into collaborative groups and their feelings or opinions regarding the collaborative work. The Authors note that this may also extend to feelings or opinions about people but this was not explored in any of our work.

- Email, reports, and models may be used to derive management information relating to process and product dependencies. This is because models are a direct representation of the product or a part thereof, reports describe technical discussion of the product, and emails describe aspects of both the product and process. Given the potential intersection and non-intersection between management information derived from the different types of asset, maximum interpretive power might be achieved through triangulation (compare and contrast) between management information derived from different types of asset.

- Similar to dependencies, management information about the level of development of the product can also be derived from email, reports and models. This is because models embody the technical definition of the product or system being engineered, and correspondingly, either implicitly or explicitly reflect the stage of design (maturity) such as definition of geometry or generation of tool paths for manufacture. Reports describe the completion of particular stages of the process including, for example, concept selection, design reviews, and testing. Lastly, email will be used to share, review and discuss models and reports, and may thus also indirectly describe progress through lexicons associated with particular types of model, reports, and discussion of their content. As with dependencies,opportunities and challenges exist

		Types of Digital Asset					
		E-Mail	Reports/ Presentations	Models	Project documents	All digital files	
Class of Management Information	No. of methods	14	7	5	4	2	Interpretation
Communications	4	✓					Project states, roles, relationships and management styles, and level of management control
Conformance	3	✓	✓	✓			Attention given to constraints, drivers and procedures
Dependencies	2	✓	✓	✓			Potential interrelationships in the process and product
Effort	2		✓	✓		✓	Types of engineering work and similarity with past projects
Product	2	✓	✓	✓	✓		Status of development of product, similarity to past products and complex product interfaces/system dependencies
Skills and competencies	4	✓	✓		✓	✓	Composition of project work, skills and knowledge of the team and sentiment (feeling) of the team
Time	2		✓	✓			Prediction of time-to-complete engineering activities
Topic	5	✓					Potential issues, level of awareness across the team and level of attention
Workflow	3		✓		✓		Identify abnormal workflow and changes that influence project complexity

Table 3: Classes of management information by types of digital asset

in triangulation between information derived from different types of asset.

- Management information concerning time and predictions thereof are derived from assets that represent the product and / or a specific activity necessary for product realisation. Correspondingly, reports and models are the primary classes of digital asset employed as the evolution or maturity of these assets directly reflects the stage or maturity of engineering. Further, given the lag between completion of work and reporting of work, models may provide real-time and leading information (when compared to traditional reporting cycles).

- Management information about overall health and normality of a project is derived from high level workflow data, which is in turn derived from project documentation and reports associated with stage-gates or formalised processes.

From this appraisal of interpretive power it is possible to gain some insights into the scope of management information that can be generated from the digital footprint and the interpretative power within the context of an engineering project.

5. Challenges, limitations and opportunities for automatically generating management information from the digital footprint

The previous sections have presented a reference model for the generation of management information from digital assets, and an appraisal of the interpretive power of the management information generated. In this section we consider the challenges, limitations, and opportunities for automatically generating management information from the digital footprint.

5.1. Scope and coverage of methods

In Section 4, a catalogue of twenty-seven distinct methods and types of management information are developed. While the twenty-seven methods and management information are not exhaustive we contend that they represent the major components of the scope of distinct, practical, and observable management information that can be derived from the digital footprint of an engineering project. Further, the methods have been derived to complement extant tool-sets and address their deficiency in provision of detailed (richer) management information relating to the state, issues and outcomes of activities that have, and are currently being undertaken. At an aggregate level the methods can provide information relating to: the content of communications; conformance of processes/practices; dependencies within a product and process; engineering effort expended; emerging and acquired skills and competencies; estimated time-to-completion of engineering; the relative attention given to topics/foci; and, analysis of workflow. In addition to the categorisation, an appraisal of interpretive power reveals that it is possible to generate management information concerning: the identification of potential issues and holistic understanding; elicitation of previously hidden process and product dependencies; assessment of the level of development of the product; real-time analysis of time expended and predictions of time remaining; and assessment of the health of a project and its normality. In order to generate this range of management information it is necessary to analyse all classes of digital asset (email, models, reports and project documentation). Further, in a number of cases management information can potentially be generated from multiple types of asset thereby enabling comparison and contrasting from different sources (types of digital asset). The correlation between type of digital asset and interpretive power is depicted in Figure 4 which highlights the one-to-one relations between email and identification of potential issues, and project health (norms) and project documentation. Figure 4 also reveals the role of reports (technical) and models in elicitation of dependencies, assessment of product development (maturity), and analysis of time spent and remaining.

5.2. Generalisability of methods

Generalisability - both within the engineering domain and beyond - depends largely on the presence of the type of digital asset. For example, the techniques employed on communications can be applied to any collaborative activity involving email. In contrast, where engineering specific tools such as CAD are employed, methods may not be applicable beyond the domain. However, it may be that the principles can be adapted to domain-specific representations (for example, digital content creation such as graphics, media and videos). Further, the methods developed relating to CAD could be applicable to other engineering modelling and analysis tools, such as, finite element (FE) analysis and computational fluid dynamics (CFD) but may require tailoring for the particular modelling tool.

In addition to domain-specific representations, domain knowledge/history is required in order to provide the basis benchmarking/comparison of project norms. For engineering, there exists a generalizable lexicon of project terms and accepted project process models (c.f. BS7000 [42] and ISO for Systems Engineering [9]) that can be employed

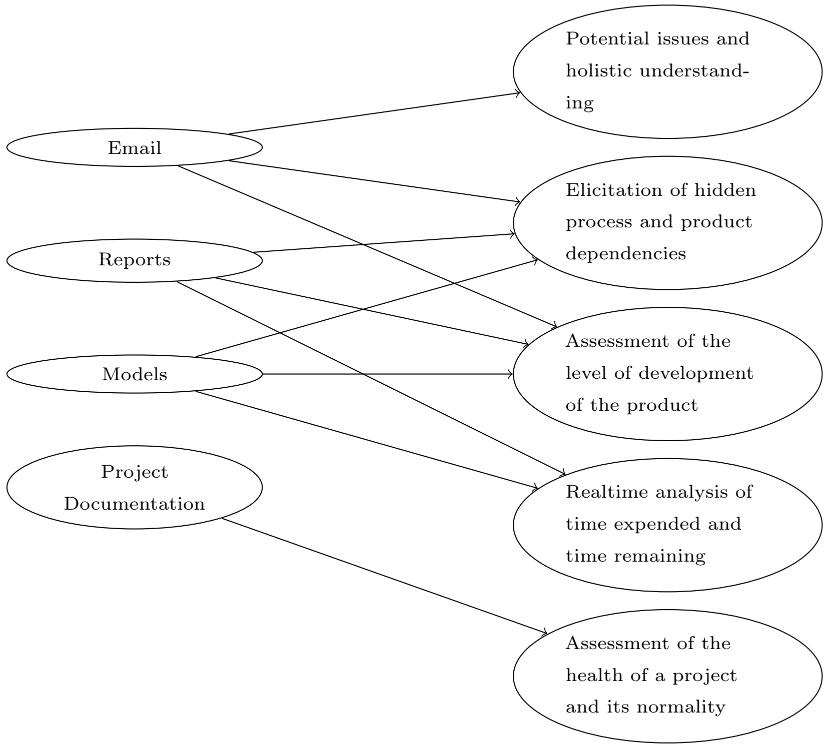


Figure 4: Types of digital asset and interpretive power

to compare and contrast collections of digital assets and their evolution during a project. This frame of reference enables the construction of data sets on which statistical analysis can be performed. Without the existence of standard models such analysis poses an almost intractable task, particularly when there currently exist no 'physics-based' models for project activity, such as those employed in IVHM [22] methods.

During the undertaking of the studies detailed in Tables 4 and 5, one of the most significant challenges for the Authors has been the verification and validation of the methods and management information. To date, we have focused primarily on whether or not the information is a fair reflection of the project and, in particular, the state, issues and outcomes of the work undertaken – i.e. no false-positives. Assessing the potential value of supplementary information for project management is a complex topic and one that we are exploring via A-B studies with and without the information [31]. Early findings suggest that the management information can encourage a much richer discussion, with a greater number of high value statements by users, where high value is deemed to be those statements that are based on evidence. Our research has shown that the information provided is not easily available via other sources, and supplements existing tools. It has also revealed the potential utility of the information provided in terms of both the technical and managerial review/insights and as an aid to understanding [31]. As with all data/information, the additional information - while potentially useful - carries an overhead in terms of processing. That is, information must be interpreted in combination with knowledge of the given context and situation [16].

5.3. Implementation of methods

One of the benefits of analysing the digital footprint (shadow) is that existing data assets are used rather than requiring the generation of additional data. However, the approach does require that the evolution of the digital footprint is captured and recorded, which while not demanding additional data does require considerably more storage than that which is required to store the digital footprint - for the studies given in this paper, typically demanding two or more orders of magnitude greater storage than the digital footprint itself [13]. The total file storage requirement is a function of the duration of a project and the monitoring intervals. In cases where the content of the digital assets is not required meta-data only can be recorded, eliminating storage issues but still requiring network access.

In terms of automation, a significant challenge lies in fully automating the data-information processing steps (Figure 3) - for the range of methods given in Tables 4 and 5 only a third were fully automated. A third (9) required limited manual input to, for example, set priorities or thresholds, while another third (8) necessitated significant manual input and interpretation, particularly where 'typing' of linguistics is necessary. The challenges of user-in-the-loop analysis are not unique to the work reported in this paper. Significant research effort is being applied to their resolution within the fields of computer science and computational linguistics where computational techniques for natural language processing to support comprehension and interaction have been heavily researched since the 1990s.

In addition to storage and automation, one of the most significant barriers to the application of automated monitoring of the digital footprint concerns privacy and monitoring of individuals. This includes but is not limited to the need to capture, access, and analyse the content of email, which continues to be a highly contentious subject with implications for potential infringement of human rights [10]). Further, the monitoring of individual work has been shown to be of concern to employees [46] who fear data may not be representative of their duties and could be used explicitly in performance management and /or could implicitly impact on promotions and rewards. While such issues have not been addressed in this work, their consideration in future work is essential if the digital footprint is to be leveraged to provide potentially important supplementary management information.

Lastly, the aim of the catalogued methods is to provide important supplementary management information, which enables project managers to be more evidence-based, in terms of the project status, and drill-down and roll-up through data representing the state, issues, and outcomes of the work that is actually being undertaken. In addition, the real-time nature of the management information provides the capability to introduce feedback and control loops, particularly regarding interventions made by the project manager/management team. That is, the impact of interventions can be evaluated with respect to the management information generated by the methods. This gives rise to the challenge of how best to present the management information to project managers – i.e. the form of visualisation and its interaction or interrogation. While not covered in this paper, recent work by the Authors has begun investigating user interface design and how best to represent the management information for the purpose of activities such as project review and management training [23].

5.4. Future outlook

Over the last decade technologies such as cloud computing, artificial intelligence and high-throughput computing have evolved to the point where many software vendors are moving to Software-as-a-Service (SaaS) and Infrastructure-as-a-Service (IaaS). The consequence of this is that the digital footprint of a project will, in the coming years, not only be created automatically as part of the engineering process, but in contrast to the studies reported in this paper, it will generated and stored in the Cloud enabling unprecedented access to its content. See for example Autodesk's Fusion 360 Cloud Platform [17]. Correspondingly, and in accord with many emerging data science industries, there is an opportunity for the development of Business Intelligence tools/Management Information Systems to exploit these assets. Such tools would need to be underpinned by scientific research aimed at characterising the cost-benefit and the potential of the tools to inform project management. Having conducted the studies reported in this paper, the Authors observed two opportunities: the potential for advanced functionality within product data management / lifecycle management systems as they migrate to the Cloud [25, 40]; and, opportunities for a new generation of real-time workflow support for not only complex mechanical, construction and systems engineering projects, but also the creative industries including media, design and computer games where products and content are created by large teams distributed across the globe.

6. Conclusion

This paper contributes to one of the major challenges present in the management of large complex engineering projects - the deficiency of current tool-sets in the provision of detailed management information that represents the state, issues and outcomes, and provide such in a cost effective manner. To remedy this deficiency the Authors have, over the past four years, undertaken a range of studies to investigate the understanding and insights that can be generated from the evolving digital footprint of an engineering project. An important tenet of the approach is that it aims to provide supplementary (complementary) real-time management information to extant tool-sets that focus on aspects of cost, quality and time, thereby enabling project managers to be more fully informed about the status of a project and, importantly, to be able to observe the impact of any interventions they make. In contrast to existing work that reports the development, verification and application of individual methods, the contribution of this paper is to review and catalogue all of the reported methods in order to appraise the scope of management information that can be automatically generated from the digital footprint. To achieve this a reference model is elicited comprised of five-stages: data acquisition, data extraction, typing and tracking, visualistaion and interpretation.

Using the reference model a set of twenty-seven methods are catalogued covering management information relating to: communications, conformance, dependencies, engineering effort, product, project, time, topic and workflow. A secondary analysis of the methods and their interpretations reveals that: email is the primary asset necessary for management information concerning potential issues and holistic understanding of the team; management information relating to process and product dependencies can be derived from multiple classes of digital asset: email, reports and models; and similarly management information about the level of development of the product can be derived from all classes of digital asset. In contrast, management information about the project health and what is referred to as normality is almost exclusively derived from project documentation which will either discuss or fulfil project deliverables. Lastly, management information concerning time and predictions thereof are derived from assets that represent the product and / or a specific activity necessary for product realisation, and correspondingly reports and models are the primary assets employed.

Following the secondary analysis the paper reflects on the scope and coverage of the set of methods and the management information they generate; their generalizability beyond the engineering domain; implementation issues; and, the future outlook. Implementation issues include practical considerations such as data capture and storage, privacy, user interface design (visualising the data) and verification and validation of the utility (cost-benefit). Lastly, the paper considers on enabling technologies such as Cloud and software-as-a-service (SaaS) contending that in the next decade the digital footprint will become more accessible and the open architectures of Cloud solutions will provide the infrastructure for the provision and integration of automated methods such as those reported in this paper.

References

[1] *APM Body of Knowledge 6th edition*. Association for Project Management, 2012. ISBN 9781903494417.

[2] *PMBOK® A Guide to the Project Management Body of Knowledge – Sixth Edition*. Project Management Institute, 2017. ISBN 9781628253825.

[3] Saeema Ahmed, Ken M. Wallace, and Luciënne T. Blessing. Understanding the differences between how novice and experienced designers approach design tasks. *Research in Engineering Design*, 14(1):1–11, Feb 2003. ISSN 1435-6066. doi: 10.1007/s00163-002-0023-z. URL https://doi.org/10.1007/s00163-002-0023-z.

[4] John Bergey, Dennis Smith, Scott Tilley, Nelson Weiderman, and Steven Woods. Why reengineering projects fail. Technical report, apr 1999. URL https://doi.org/10.21236%2Fada362725.

[5] CADArtifex. *AutoCAD 2018: A Power Guide for Beginners and Intermediate Users*. CreateSpace Independent Publishing Platform, 2018. ISBN 10: 1547211911.

[6] Alton Chua and Wing Lam. Why KM projects fail: a multi-case analysis. *Journal of Knowledge Management*, 9(3):6–17, jun 2005. doi: 10.1108/13673270510602737. URL https://doi.org/10.1108%2F13673270510602737.

[7] Ruth Deakin Crick, Cristina Stringher, and Kai Ren, editors. *Learning to Learn*. Routledge, apr 2014. doi: 10.4324/9780203078044. URL https://doi.org/10.4324%2F9780203078044.

[8] Anton De Wit. Measurement of project success. *International journal of project management*, 6(3):164–170, 1988.

[9] International Organisation for Standards. *ISO/IEC/IEEE 15288:2015 Systems and software engineering – System life cycle processes*. International Organisation for Standards, 2015.

[10] GDR. Echr rules against historic uk bulk interception regime, September 2018. URL https://globaldatareview.com/article/1174183/echr-rules-against-historic-uk-bulk-interception-regime.

[11] J.A. Gopsill, L. Shi, C. McMahon, and B.J. Hicks. The engineering design process through an understanding of the evolution of engineering digital objects. In *DESIGN 2014*, 2014.

[12] James Gopsill, Hamish McAlpine, and Ben Hicks. Supporting engineering design communication through a social media tool - insights for engineering project management. Proceedings of DESIGN 2014, 13th International Design Conference, pages 1785–1798, 2014.

[13] James Gopsill, Chris Snider, Lia Emanuel, Sian Joel-Edgar, and Ben Hicks. Automatic design structure matrices: A comparison of two formula student projects. 2017.

[14] James A. Gopsill, Chris Snider, Chris McMahon, and Ben Hicks. Automatic generation of design structure matrices through the evolution of product models. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 30(4):424–445, 2016. doi: 10.1017/S0890060416000391.

[15] B. J. Hicks, A. Dong, R. Palmer, and H. C. Mcalpine. Organizing and managing personal electronic files. *ACM Transactions on Information Systems*, 26(4):1–40, sep 2008. doi: 10.1145/1402256.1402262. URL https://doi.org/10.1145%2F1402256.1402262.

[16] B.J. Hicks, S.J. Culley, R.D. Allen, and G. Mullineux. A framework for the requirements of capturing, storing and reusing information and knowledge in engineering design. *International Journal of Information Management*, 22(4):263 – 280, 2002. ISSN 0268-4012. doi: https://doi.org/10.1016/S0268-4012(02)00012-9.

[17] Autodesk Inc. Fusion 360: Integrated cad, cam, and cae software in the cloud, 2019. URL https://www.autodesk.com/products/fusion-360/overview.

[18] Liquidplanner Inc. Liquidplanner: online project management solution, 2018. URL https://www.liquidplanner.com/features/.

[19] Mavenlink Inc. Mavenlink: Project management software & tools, 2018. URL https://www.mavenlink.com/tour/project-management-tools.

[20] ISO. Iso 21500:2012 guidance on project management, 2012. URL https://www.iso.org/standard/50003.html.

[21] Lei Shi James Gopsill, Chris Snider and Ben Hicks. Modelling the evolution of computer aided design models: Investigating the potential for supporting engineering project management. In *Product Lifecycle Management in the Era of Internet of Things*, IFIP Advances in Information and Communication Technology, pages 344–354. Springer International Publishing, 4 2016. ISBN 9783319331102. doi: 10.1007/978-3-319-33111-9_32.

[22] Ian K Jennions, editor. *Integrated Vehicle Health Management: The Technology*. SAE International, sep 2013. doi: 10.4271/r-429. URL https://doi.org/10.4271%2Fr-429.

[23] S. Joel-Edgar and J. Gopsill. Understanding user requirements in context: A case study of developing a visualisation tool to map skills in an engineering organisation. In *2018 International Conference on Information Management and Processing (ICIMP)*, pages 6–10, Jan 2018. doi: 10.1109/ICIMP1.2018.8325832.

[24] Sian Joel-Edgar and James A. Gopsill. Understanding user requirements in context: A case study of developing a visualisation tool to map skills in an engineering organisation. In *Proceedings of the International Conference on Information and Processing*, 2017. URL https://jamesgopsill.github.io/Publications/papers/conference/icimp2018/icimp2018.pdf.

[25] David Edward Jones, Chris Snider, Jason Yon, James Gopsill,

Yifan Xie, Nicolas Chanchevrier, and Ben Hicks. The design of visual information objects in three-dimensional virtual environments for engineering information navigation. Proceedings of the DESIGN 2018 15th International Design Conference.

[26] H Kundnani. What's the problem with airbus? 2006. URL https://www.theguardian.com/business/2006/oct/05/theairlineindustry.travelnews.

[27] Thomas Kvan. Collaborative design: what is it? *Automation in Construction*, 9(4):409 – 415, 2000. ISSN 0926-5805. doi: https://doi.org/10.1016/S0926-5805(99)00025-4. URL http://www.sciencedirect.com/science/article/pii/S0926580599000254.

[28] Stan Lipovetsky, Asher Tishler, Dov Dvir, and Aaron Shenhar. The relative importance of project success dimensions. *R and D Management*, 27(2):97–106, apr 1997. doi: 10.1111/1467-9310.00047. URL https://doi.org/10.1111%2F1467-9310.00047.

[29] Hamish McAlpine, Ben J Hicks, Greg Huet, and Steve J Culley. An investigation into the use and content of the engineer's logbook. *Design Studies*, 27(4):481–504, 2006.

[30] MIMOSA. Open system architecture for condition-based maintenance, 2019. URL http://www.mimosa.org/mimosa-osa-cbm/.

[31] N Mogles, L Emanuel, J Gopsill, S Joel-Edgar, K Robinson, B Hicks, D Jones, and L Newnes. Information visualisation for project management: Case study of bath formula student project. DCC'18: Design, Computing & Cognition, 2018.

[32] M. Morisio, M. Ezran, and C. Tully. Comments on "more success and failure factors in software reuse". *IEEE Transactions on Software Engineering*, 29(5):478, may 2003. doi: 10.1109/tse.2003.1199077. URL https://doi.org/10.1109%2Ftse.2003.1199077.

[33] Jeffrey K. Pinto and Dennis P. Slevin. Critical factors in successful project implementation. *IEEE Transactions on Engineering Management*, EM-34(1):22–27, feb 1987. doi: 10.1109/tem.1987.6498856. URL https://doi.org/10.1109%2Ftem.1987.6498856.

[34] J.K. Pinto and S.J. Mantel. The causes of project failure. *IEEE Transactions on Engineering Management*, 37(4):269–276, 1990. doi: 10.1109/17.62322. URL https://doi.org/10.1109%2F17.62322.

[35] Steven Poltrock, Jonathan Grudin, Susan Dumais, Raya Fidel, Harry Bruce, and Annelise Mark Pejtersen. Information seeking and sharing in design teams. In *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work - GROUP'03*. ACM Press, 2003. doi: 10.1145/958197.958198. URL https://doi.org/10.1145%2F958197.958198.

[36] Lei Shi, Linda Newnes, Stephen Culley, James Gopsill, Simon Jones, and Christopher Snider. Identifying and visualising kpis for collaborative engineering projects: a knowledge based approach. 7 2015. ICED15, The 20th International Conference on Engineering Design, 2015 ; Conference date: 27-07-2015 Through 31-07-2015.

[37] Lei Shi, Linda Newnes, Steve Culley, and Bruce Allen. Learning to predict characteristics for engineering service projects. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 31(3):313–326, 2017. doi: 10.1017/S0890060416000470.

[38] C. Snider, J.A. Gopsill, S. Jones, L. Shi, and B. Hicks. Understanding engineering projects: An integrated vehicle health management approach to engineering project monitoring. In *Proceedings of the International Conference on Engineering Design, ICED*, volume 3, 2015.

[39] Chris Snider, Lia Emanuel, James Gopsill, Sian Joel-Edgar, and Ben Hicks. Identifying the influences on performance of engineering design and development projects. DS 87-2 Proceedings of the 21st International Conference on Engineering Design (ICED 17), pages 299–308, 2017. ISBN 978-1-904670-90-2.

[40] Chris Snider, James Gopsill, David Jones, and Ben Hicks. PLM 2018: Product Lifecycle Management, 2018.

[41] Chris Snider, James A. Gopsill, Simon L. Jones, Lia Emanuel, and Ben J. Hicks. Engineering project health management: A computational approach for project management support through analytics of digital engineering activity. *IEEE Transactions on Engineering Management*, pages 1–12, 2018. doi: 10.1109/tem.2018.2846400. URL https://doi.org/10.1109%2Ftem.2018.2846400.

[42] British Standards. *BS 7000 Series - Design Management Systems*. British Standards, 1996.

[43] R Syal. Abandoned nhs it system has cost £10bn so far. 2013. URL https://www.theguardian.com/society/2013/sep/18/nhs-records-system-10bn.

[44] Mohammadreza Tahan, Elias Tsoutsanis, Masdi Muhammad, and Z.A. Abdul Karim. Performance-based health monitoring, diagnostics and prognostics for condition-based maintenance of gas turbines: A review. *Applied Energy*, 198:122 – 144, 2017. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2017.04.048.

[45] Telegraph. Boeing 787 dreamliner: a timeline of problems. 2013. URL https://www.telegraph.co.uk/travel/comment/Boeing-787-Dreamliner-a-timeline-of-problems/.

[46] David L. Tomczak, Lauren A. Lanzo, and Herman Aguinis. Evidence-based recommendations for employee performance monitoring. *Business Horizons*, 61(2):251 – 259, 2018. ISSN 0007-6813. doi: https://doi.org/10.1016/j.bushor.2017.11.006.

[47] J. Wasiak, B. Hicks, L. Newnes, C. Loftus, A. Dong, and L. Burrow. Managing by e-mail: What e-mail can do for engineering project management. *IEEE Transactions on Engineering Management*, 58(3):445–456, Aug 2011. ISSN 0018-9391. doi:

10.1109/TEM.2010.2090160.

[48] James Wasiak, Ben Hicks, Linda Newnes, Andy Dong, and Laurie Burrow. Understanding engineering email: the development of a taxonomy for identifying and classifying engineering work. *Research in Engineering Design*, 21(1):43, Aug 2009. ISSN 1435-6066. doi: 10.1007/s00163-009-0075-4. URL https://doi.org/10.1007/s00163-009-0075-4.

[49] E. Westerveld. The project excellence model®: linking success criteria and critical success factors. *International Journal of Project Management*, 21(6):411–418, aug 2003. doi: 10.1016/s0263-7863(02)00112-6. URL https://doi.org/10.1016%2Fs0263-7863%2802%2900112-6.

[50] Wikipedia. Sydney opera house. 2018. URL https://en.wikipedia.org/wiki/Sydney_Opera_House.

[51] Wikipedia. Olkiluoto nuclear power plant. 2018. URL https://en.wikipedia.org/wiki/Olkiluoto_Nuclear_Power_Plant.

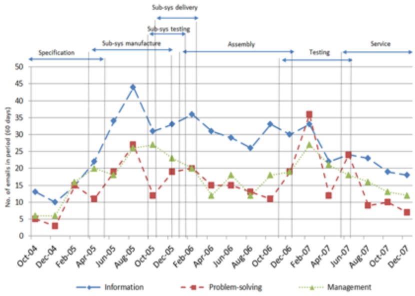
Appendix A. Catalogue of methods for generating management information from the digital footprint of engineering projects

Class	Method No.	Method Description	Data Source	Extracted Data	Typing & Tracking	Interpretation
Communication	M1	Typing of communications by management purpose: info sharing, problem solving and management activity	E-mail / Social Media	E-mail: time sent and body content (thread, subject and to, cc, from excluded)	Typing and tracking of relative level of occurrence of emails typed by purpose over time (manually typed).	Eight project states can be established from the combinations of relative change in the three traces. State include: pressure points, steady workings, working nearing completion, completed work, management input required to co-ordinate work, information required to continue work, information sufficient but management input needed to control and information and control sufficient.
Communication	M2	Typing of communications by management purpose: product, process and people	E-mail / Social Media	E-mail: time sent and body content (thread, subject and to, cc, from excluded)	Typing and tracking of relative level of occurrence of emails typed by subject over time (manually typed).	Relative changes in focus of work and management on product, people and process.
Communication	M3	Monitoring patterns of exchange within the communication networks of projects	E-mail / Social Media	E-mail: time sent, sender and recipient(s)	Determination of the overall and relative levels of communication within the network and between individuals and groups within the network.	Identification of gatekeepers, indication of spread of communication and insights into relationships (connections).
Communication	M4	Monitoring the composition of community by contribution to the communications network (email style)	E-mail / Social media	E-mail: time sent, sender and recipient(s), subject and length of body content	Clustering email style for member of the community by no. of emails sent; no. of recipients, length of subject and length of body content.	Indication of potential issues through sudden changes in composition (styles); profiling of contributors by style; indication of composition for a 'healthy community' i.e. no. of detailed responders.
Conformance	M5	Monitoring the relative attention given to the product/project requirements	Email, reports and presentations	Documents & email: time, title and content	Tracking the relative and cumulative levels of occurrence (utterances) of terms from the requirements that appear within the project documentation and emails over time.	Identify terms from the requirements that have received the greatest and lowest (no) attention.
Conformance	M6	Monitoring the relative attention given to relevant regulations and/or standards	Email, reports and presentations	Documents & email: time, title and content	Tracking the relative and cumulative levels of occurrence (utterances) of terms from relevant regulations/standards/legislation that appear within the project documentation and emails over time.	Identify terms from regulations/standards/legislation that have received the greatest and lowest (no) attention.
Conformance	M7	Assessment of conformance of content to standard operating procedures	Models, reports, CAD files and code	Files: time, title and content	Extraction and assessment of the content of documents to best practice/standards for construction/execution/ Including document structure, CAD model structure and parameter values.	Monitor compliance of project/engineering work. Highlight non-conformance.
Dependencies	M8	Eliciting potential project dependencies through co-occurrence of modifications to types of digital file	Models, reports, CAD files and code	Files: type, size, access date and last saved date	Revealing project, process and product dependencies by co-occurrence and clustering of modifications to typed files, such as CAD.	Identify potentially hidden or emerging dependencies between physical parts (CAD, simulation) and tasks, activities or deliverables (reports, presentations).
Dependencies	M9	Monitoring associations and interconnections between areas of project work	Emails, reports and presentations	email: time sent and subject	Characterisation (typing) of topics clustered by co-occurrence over time using strategic diagrams to represent relation between centrality and density: emerging or declining; basic and traversal; developed and isolated; and developed and core.	Insights into work complexity, breadth of focus, core topics, isolated topics, coherence of topics, and divergence/convergence. Major changes indicate a shift in focus and potential transition.
Engineering effort	M10	Evaluate the level of reuse of content from previous projects	Models, reports, CAD files and code	Files: time, title and content	Extraction of content of project files and comparison to the content of similar file types from previous projects. Applicable to CAD models, documents and models.	Monitor the similarity (re-use) of project/engineering work from past projects. Also, identify potential novelty/new design/ideas/approaches particularly if files reduce in similarity.
Engineering effort	M11	Assessment of the type and distribution of engineering work	All digital assets	Files: type, size, access date and last saved date	Tracking of the relative access and modification to digital files that are typed by engineering work activities such as concept design, detailing, manufacture.	Monitoring of the type of engineering work undertaken by individuals and the project team.
Product	M12	Assessment of the status (level of development of a design)	CAD files	CAD files: content; access date and last saved date	Analysis of CAD files with respect to the inclusion of CAD functions that are typed with respect to the level of development of design tasks.	Assessment of the level of development of a design and stage of the design process, such as tool path generation.
Product	M13	Elicitation of the product architecture/structure from project communication and documentation	Email, reports and presentations	Documents & email: time; title and content	Co-occurrence and clustering of product-related terms extracted through term-frequency inverse document-frequency.	Reveal evolving product architecture/structure 'as is' being designed (receiving attention/effort).
Skills & Competencies	M14	Assessing the similarity of projects through comparison of the content of project briefs	Project documentation	Documentation: project title and brief (request)	Similarity of content of project brief to past project briefs to identify the closest matching past projects.	Identification of similar/typical projects that may be more routine and that have predictable duration and can therefore be planned and resourced with more confidence.
Skills & Competencies	M15	Elicitation of the development of knowledge and competencies within a team or organisation	Technical reports	Technical reports: date, authors and textual content	Analysis of terms and their co-occurrence followed by clustering and network analysis of terms used by authors within a document corpus over time.	Indication of the expertise of individuals and/or the emerging or changing knowledge and competencies of an individual, group or organisation.
Skills & Competencies	M16	Assessment of the sentiment, affect and tone of project members	Email/social media	Email: Time sent; sender and body content	Application of sentiment analysis tools to score the sentiment of the content of emails of project team.	Identify individuals with changing or particularly string sentiment.
Skills & Competencies	M17	Assessment of the type and distribution of project work	All digital assets	Files: type; size; access date and last saved date	Tracking the relative level and distribution of attention to classes of digital asset typed by project function, such as reporting, planning, risk register and management.	Monitoring of the type of project work undertaken by individuals and the project team relative to stage-gates.

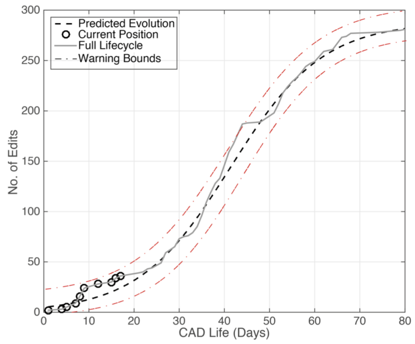
Table 4: Methods for generating management information from the digital footprint of engineering projects

Class	Method No.	Method Description	Data Source	Extracted Data	Typing & Tracking	Interpretation
Time to completion	M18	Predicting the time to complete a task through the rate of modification of a typed file	Models, reports, CAD files and code	files: type; size; access date and last saved date	Cumulative modifications to a file type compared with functions derived from previous cumulative edits to type of file e.g. CAD file exhibits a Sigmoid function.	Assess the relative completeness of tasks and the predicted time to complete.
Time to completion	M19	Predicting the time to complete activity or stage through the occurrence of typed events	Models, reports, CAD files and code	Files: type	Comparison of relative occurrence of typed events with respect to historical events.	Assess the relative completeness of an activity and over/undershoot.
Topic	M20	Tracking and typing topics in communication by diffusion characteristics	Email	Emails: to/cc/from; time sent; body content (thread and subject excluded)	Typing of topics by frequency and patterns of occurrence of terms relative to each other over time. Diffusion types include: short-duration normal-activity; long-duration normal-activity; long-duration normal-activity high-initial membership; high relevance; high intensity short-duration activity; and, high transmission low-spread activity.	Identify topics exhibiting abnormal diffusion characteristics; highlight topics that may represent potential issues; contrast diffusion of topics with expectations; establish levels of topic awareness across the network and identify key individuals.
Topic	M21	Tracking the evolution of clusters of topics in email communications	Email	Emails: time sent; subject or body content	Continuous clustering of topics and tracking of the relative occurrence of clusters over time.	Judge the breadth, divergence and convergence of ongoing work and changes in focus; identify potential issues and outliers. Particularly important around stage gates.
Topic	M22	Tracking the relative levels of attention of an individual or group to a topic	Email	Emails: to/cc/from; time sent; email body content (thread and subject excluded)	Identification and tracking of the frequency of occurrence of a topic (term) by an individual or group.	Indication of the work focus of an individual or group; identify the issues being dealt with by an individual or group - particularly unresolved issues. Provide insights into decision-makers.
Topic	M23	Assessment of the sentiment, affect and tone of members with regard to a project topic	Email/social media	Email: time sent; sender; body content	Application of sentiment analysis to score the sentiment of the content of emails relating to project related terms. Project related terms extracted through term frequency inverse document frequency of the evolving email corpa.	Identify project related terms (topics) with strongly changing or particularly strong sentiment.
Topic	M24	Assessment of the spread and diffusion of topics within the project team	Email/social media	Email: time sent; sender and body content	Frequency and patterns of topics in communication over time and within project teams. Measures include spread, speed, intensity, shock and persistence of topics.	Identify people/teams who know about topics (awareness). Highlight topics that spread rapidly and/or are persistent - potentially unresolved or requiring management intervention.
Workflow	M25	Assessing the similarity of projects through comparison of the sequences of workflow in past projects	Project documentation	documentation: time and type	Documents typed by purpose or function, and the evolving sequence of occurrence of documents compared to document sequences in past projects. Requires taxonomy of document types.	Identification of similar / typical project phases that may be more routine and that have predictable duration and can therefore be planned and resourced with more confidence.
Workflow	M26	Assessing the normality of a project's workflow through comparison with past projects	Project documentation	documentation: time and type	Documents typed by purpose or function, and the evolving sequence of occurrence of documents compared to the most commonly occurring sequences of documents in past projects.	Indication of the projects that are or have become outliers (atypical) and require management attention/review.
Workflow	M27	Assessing a projects' level of complexity through comparison of workflow with previously categorised projects	Project documentation	documentation: time and type	Documents typed by purpose or function, and the evolving sequence of occurrence of documents compared to sequences of occurrence of documents from past project that are clustered by level of complexity and compared to the most commonly occurring sequences of documents in past projects.	Indication of the likely complexity and/or change in complexity of a project. Informs potential changes in duration, difficulty and resourcing.

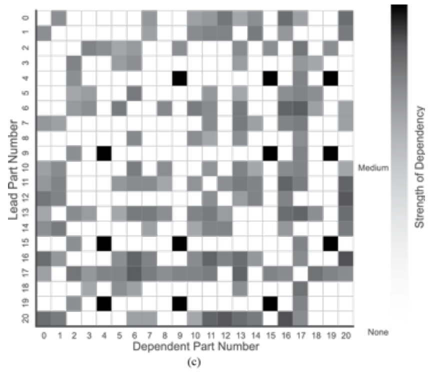
Table 5: Methods for generating management information from the digital footprint of engineering projects cont.



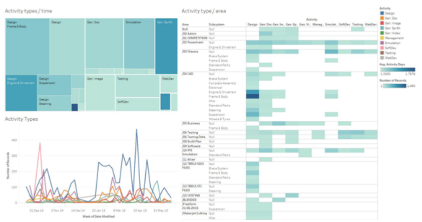
(a) Example communication information visualisation: Time-series trace showing relative levels of different communication types by month for a single project [47].



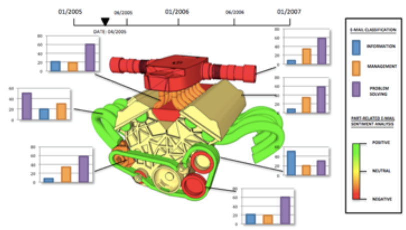
(b) Example conformance information visualisation: Sigmoid functions as a means to characterise normal development profiles [21].



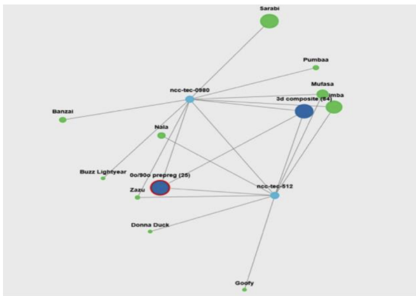
(c) Example dependency information visualisation: Adjacency matrix showing inter-file dependencies and likelihood of impact from change. [13].



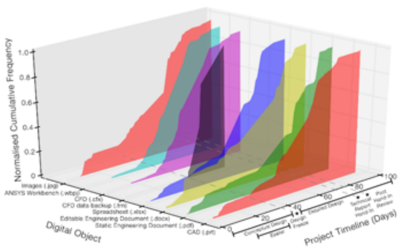
(d) Example Engineering Effort information visualisation: Table view and area chart showing effort with respect to areas of the system and engineering activity [31].



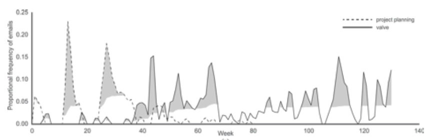
(e) Example Product information visualisation: Heat map of sentiment against components [25].



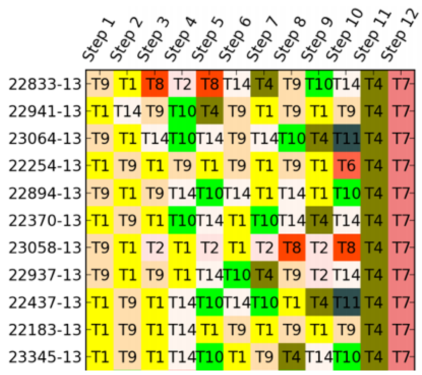
(f) Example Skills & Competencies information visualisation: A network showing the relative occurrence of terms and their use by engineers [24].



(g) Example Time information visualisation: Stacked area charts of cumulative use of digital tools through file access [11].



(h) Example Topic information visualisation: Area chart showing relative attention to topics by project team over time [41].



(i) Example Workflow information visualisation: Matrix showing sequences of workflow for projects of similar levels of complexity [36].

Figure 5: Part (a - i) - Example visualisations from each class of management information